

Segmentación de clientes y definición de alertas para la prevención de riesgos de lavado de activos y financiación del terrorismo (SARLAFT): un estudio económico aplicado a entidad financiera colombiana en 2017.

Estudiante:

Mateo Amaya Molina

Asesor:

Juan Camilo Chaparro Cardona

Universidad EAFIT

Escuela de Economía y Finanzas

Medellín

2017

Resumen: Los gobiernos del siglo XXI tienen el reto de fortalecer su estructura y controles sobre los crímenes financieros de los cuales pueden ser objeto. El lavado de activos y la financiación del terrorismo son los crímenes que se cometen con más frecuencia alrededor de todo el mundo. Este tipo de actividades vulnera el equilibrio económico de una manera sigilosa y a su vez constituyen un riesgo a la hora de hablar de relaciones comerciales internacionales. Los avances informáticos en campos como la minería de datos y la estadística, han creado un panorama completamente nuevo para los gobiernos que pretenden generar controles para prevenir y detectar este tipo de actividades. Estos controles son exigidos a las diferentes entidades privadas que operan en el país, según el riesgo que represente para el sector de la economía al que pertenecen. El presente escrito desarrolla una propuesta metodológica conformada por la ejecución de técnicas de minería de datos y el análisis económico sectorial, para dar una posible respuesta al control diseñado por el regulador colombiano de las entidades del sector financiero.

Abstract: The XXI century governments get the challenge to fortress their internal structure and controls against financial crimes of which they can be object. The money laundering and the terrorism financing are the most frequently committed financial crimes around the world. This type of activities is threatening the economic equilibrium in silence way turn and it's representing a risk while talking about international trade relationships too. The informatics advances in the field of data mining and statistics created a new complete landscape for those governments that try to get preventing controls and to identify that kind of activities. These controls are required to those type of private organizations operating in the country in depending of the precise risk to each economic sector they belong. This research develops an innovative method composed by the combination of techniques of data mining applying and the economics sector analysis looking for a possible answer to the financial sector entities controlling designed by SARLAFT in the Colombian government.

Tabla de contenido

1. Introducción y planteamiento del problema.....	3
1.1 Preguntas.....	10
2. Objetivos	11
3. Justificación	11
4. Marco teórico.....	12
4.1 Diccionario Conceptual.....	12
4.2 Marco Teórico y antecedentes:	13
4.3 Planteamiento de la hipótesis.....	15
5. Metodología.....	15
5.1 Descripción de la información	15
5.1.1 Variables	15
5.1.2 Observaciones	16
5.1.3 Software implementado.....	17

5.2 Descripción de la metodología seleccionada	17
6. Resultados de la metodología de segmentación	22
6.1 Personas naturales.....	22
6.2 Personas jurídicas.....	24
6.3 Alertas personas jurídicas.....	26
7. Análisis económico sectorial de las alertas generadas.....	27
8. Conclusiones	32
Bibliografía	33
Apéndice.....	35

1. Introducción y planteamiento del problema.

Uno de los mayores retos para las economías del siglo XXI consiste en fortalecer su estructura y composición frente a fenómenos paraestatales que procuran desestabilizar un equilibrio económico y social prescrito por el Estado. Dentro de estos fenómenos se encuentran prácticas como el contrabando, las guerrillas y los diferentes crímenes financieros. Estas prácticas desde diferentes enfoques desequilibran el modelo económico adoptado por un país. Su ejecución tiene como común denominador ignorar la legitimidad del Estado y sus composiciones normativas.

Las consecuencias que un país puede enfrentar por tener el desarrollo de este tipo de actividades dentro de su territorio, van desde problemas de sostenibilidad fiscal y financiera derivados de la evasión de impuestos e ingreso de dinero derivado de actividades delictivas, hasta problemas sociales tan serios como la inseguridad, la violencia y el terrorismo. Para efectos del presente estudio se hará énfasis en el lavado de activos y la financiación del terrorismo, entendidos como crímenes financieros que afectan la estabilidad económica de Colombia.

Los delitos financieros son considerados como un conjunto de actividades realizadas por individuos u organizaciones con el propósito de obtener ventaja ilegal sobre determinadas situaciones. Su ejecución casi nunca está acompañada con fuerza o violencia, pero sí afectan a millones de personas y cuestan una gran cantidad de recursos a los Estados (Sudjinato, y otros, 2010). Para la Unidad de Información y Análisis Financiero (UIAF)¹, los delitos financieros más importantes de combatir son el lavado de activos y la financiación del terrorismo (LA/FT)², los

¹ La UIAF es la unidad administrativa especial adscrita al Ministerio de Hacienda y Crédito Público encargada de proteger la seguridad nacional en el ámbito económico, por medio de la detección, prevención y lucha contra el lavado de activos y financiación del terrorismo.

² Entiéndase esta sigla como la abreviación de lavado de activos y financiación del terrorismo.

cuales consisten en la realización de acciones que buscan dar apariencia de legalidad a los recursos generados por actividades ilícitas y apoyar de cualquier manera el terrorismo o a los grupos que lo fomentan, respectivamente (UIAF, 2014).

Los primeros intentos por controlar e identificar este tipo de actividades se remontan a mediados de la segunda mitad del siglo XX en los Estados Unidos, cuando este país comenzó a implementar medidas para evitar que sus nacionales usaran bancos internacionales para evadir impuestos. Estas medidas se tradujeron en la creación de 101 FIU's³ para antes del año 1999, encargadas de identificar y reportar actividades sospechosas al interior de ese país (Levi & Reuter, 2006).

De la mano del Banco Mundial, el FMI y las fuerzas de los diferentes países del G8, este tipo de controles se consolidó para finales de los 80's con la creación de la institución intergubernamental llamada Financial Action Task Force on Money Laundering (FAFT-GAFI). Esta institución se convertiría en la herramienta para hacer política internacional sobre el tema en todo el mundo, ya que sus recomendaciones y metodologías sobre la identificación, control y monitoreo de riesgos sobre el LA/FT debían ser cumplidas por los países miembros de la comunidad internacional por medio de la coacción de posibles sanciones, bloqueos o multas (Levi & Reuter, 2006).

Una vez conocido el origen de la institución internacional encargada de dictaminar la razón y las formas de implementar controles por parte de los países sobre el lavado de activos y la financiación del terrorismo, es necesario saber cómo éstas son acogidas y ejecutadas por el gobierno colombiano dentro del modelo económico que custodian. Para el Instituto Basel de Gobernanza⁴, Colombia es un país que ejecuta muy buenas prácticas en cuanto la prevención y monitoreo de estos crímenes financieros. Este instituto elabora anualmente un índice llamado “Basel Anti-Money Laundering” para calificar de 0 a 10 qué tan alto es el riesgo que tiene un país de ser objeto de este tipo de crímenes. Para el 2017 Colombia ocupa el puesto 125°, siendo el 1° el más riesgoso y el 149° el menos riesgoso, con una calificación de 4.57. Este puesto lo convierte en un país de bajo riesgo y el menos riesgoso de todo América por encima de países como Chile, Estados Unidos y Canadá (BAMLI, 2017).

³ Unidades de inteligencia financiera.

⁴ El instituto Basel de Gobernanza es un instituto internacional e independiente dedicado a trabajar con sectores públicos y privados para combatir la corrupción, el lavado de activos y mejorar la calidad de la gobernanza.

Los resultados que propone este índice, en principio, son problemáticos y tienden a desentonar con la percepción general que se tiene de Colombia sobre este tema. Es suficiente con mirar menos de 40 años de historia colombiana, para sustentar esta percepción negativa sobre el tema en Colombia. Ahora bien, para ver qué tan coherente puede llegar a ser el índice Basel AML, se investigó sobre la metodología que este índice aplica para calcular el riesgo LA/FT de cada uno de los países. Esta información está disponible en el “*Basel AML Index 2017 Report*”, el cual, es un reporte publicado anualmente por ese instituto y justifica la construcción del índice por medio de la ponderación realizada de 14 indicadores sobre el tema. Estos indicadores se obtienen de fuentes públicas de información y van desde la percepción de la corrupción y transparencia financieras; hasta el cumplimiento de recomendaciones de autoridades internacionales como FAFT (Basel Institute on Governance, 2017). Confirmar que la metodología de este índice tenga buenos cimientos metodológicos, deja el panorama nublado a la hora de sacar conclusiones de Colombia sobre este tema, porque su buen puntaje quizá estaría sustentado en la existencia de estrictas normas y controles frente al tema, pero ineficaces en la práctica.

Por otro lado, el gobierno nacional por medio de una Evaluación Nacional De Riesgo De Lavado De Activos Y Financiación Del Terrorismo, presenta los resultados sobre las fragilidades y amenazas que tiene el país frente a los riesgos LA/FT. Estos resultados identifican que el país en general tiene una amenaza alta de que se laven activos, puesto que para la actualidad se siguen cometiendo muchas actividades como el narcotráfico, contrabando, minería ilegal, corrupción y extorsión dentro del país (Universidad del Rosario, MinJusticia & UIAF, 2016).

Ambas posiciones generan un panorama de confusión y desconcierto en la intención de comprender cuál es la realidad sobre este tema en el país, pues si bien, el instituto Basel otorga un parte de tranquilidad desde la existencia de políticas suficientes para controlar este tipo de crímenes; el estudio realizado a nivel nacional sustenta de manera muy coherente por qué Colombia sigue con un alto riesgo de padecer este tipo de perversiones dentro de sus actividades.

Partiendo de la certeza de la existencia de regulaciones estrictas frente al tema, surge la cuestión de saber cómo está diseñado el andamiaje jurídico que rige este tema en Colombia. Si bien se mencionó que la UIAF es la delegada por el Ministerio de Hacienda para suministrar y manipular la información sobre el tema, corresponde a la Superintendencia Financiera de Colombia⁵ en su función de vigilancia y control de todas las entidades que presten servicios financieros en

⁵ Organismo técnico adscrito al Ministerio de Hacienda y Crédito Público.

Colombia, exigir el cumplimiento por parte de éstas de las normas internacionales sobre el manejo de riesgos y el lavado de activos.

Para tal efecto, la Superintendencia dentro de su Circular Básica Jurídica⁶, específicamente en la Parte I, Título IV, tiene un Capítulo específico para el cumplimiento de estas recomendaciones llamado “*Instrucciones relativas a la administración del riesgo de lavado de activos y de la financiación del terrorismo*”. Este capítulo instruye a todas las entidades vigiladas cómo deben implementar un Sistema de Administración del Riesgo de Lavado de Activos y Financiación del Terrorismo (SARLAFT). Este sistema le es exigido a todas las entidades vigiladas por la Superintendencia Financiera para poder operar en el país.

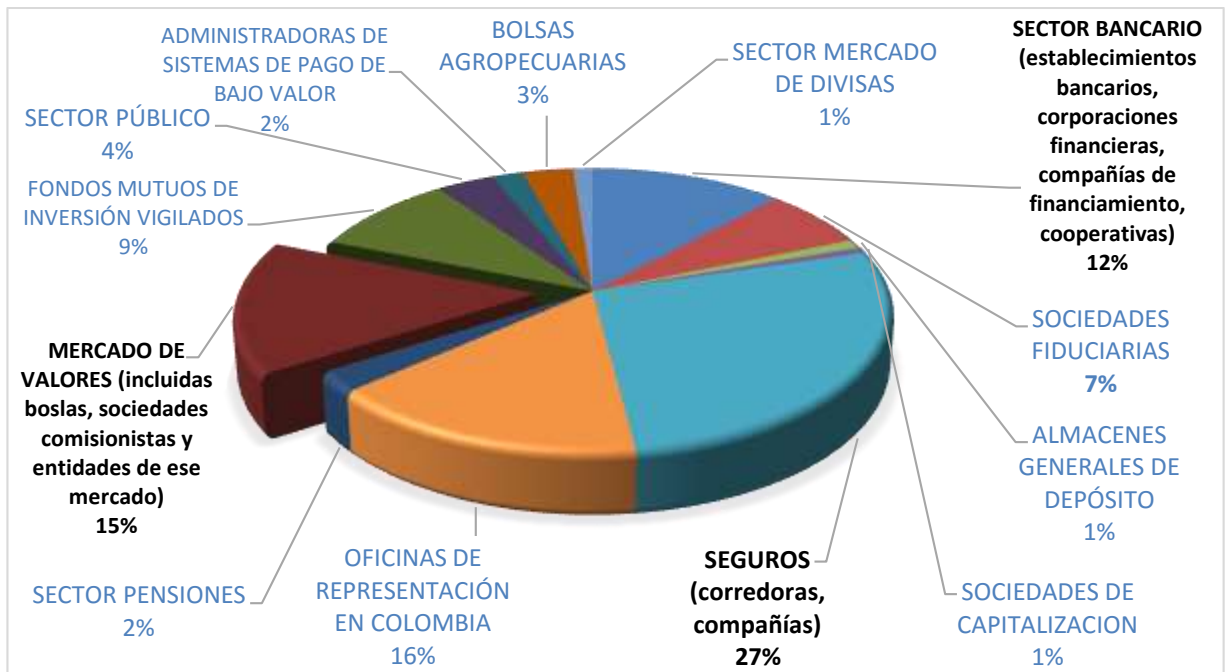
Teniendo claro en qué consisten las actividades desvaloradas⁷ judicialmente por Colombia y los diferentes países con los que puede tener relaciones comerciales, es pertinente aclarar cuáles son los diferentes instrumentos y canales por los que estas actividades se podrían materializar al interior del país. Para tener una idea del alcance que tienen las regulaciones promulgadas por la Superintendencia Financiera sobre el LA/FT, hay que mencionar que bajo su vigilancia se encuentran 417 entidades de diferentes naturalezas que acuden al mercado financiero, bien sea para prestar servicios al mismo sector o para intermediar entre los ciudadanos y el sistema financiero. En el gráfico 1 se muestra la composición de estas entidades vigiladas.

Los sectores resaltados (seguros, bancos y mercado de valores), representan la mayor amenaza por medio de la cual podrían permear dineros con orígenes ilícitos. Estos representan un 54% del total de las entidades vigiladas y son los sectores donde la Superintendencia hace mayor énfasis en la ejecución del SARLAFT. Puesto que el sector bancario y el mercado de valores cumplen la función de captar dinero y canalizarlo por medio de inversiones como títulos, acciones, CDT's o cuentas de ahorro; para la Superintendencia representan el canal más latente por medio del cual podrían cometerse los crímenes financieros objeto del presente estudio.

Gráfico 1. Sectores vigilados por la Superintendencia Financiera.

⁶ Entiéndase a la Circular Básica Jurídica como la herramienta jurídica por medio de la cual el gobierno regula todo lo atiniente a la actividad financiera en el país.

⁷ Desde una perspectiva jurídico-penal, la desvaloración de conductas o actividades, se entiende como la percepción negativa por parte del estado o la sociedad sobre una acción determinada, la cual, puede perturbar la tranquilidad normal de la comunidad, y, por lo tanto, su juzgamiento desde los preceptos del derecho penal.



Fuente: Superintendencia Financiera de Colombia, 2017. (cálculos propios).

Estas entidades financieras deben demostrar una correcta ejecución del SARLAFT, el cual está compuesto por una serie de etapas que consisten en la identificación, medición, control y monitoreo de los factores de riesgo LA/FT. La etapa de la identificación es la columna vertebral del sistema si se entiende que ésta delimita cuales son los sujetos y las transacciones potencialmente contaminadas por este tipo de crímenes.

Merece la pena aclarar que la entidad financiera objeto de este estudio hace parte de los actores que acuden al mercado de valores a prestar servicios de corretaje de valores, banca de inversión, manejo de encargos fiduciarios y administración de Fondos de Inversión Colectiva. Es una de las firmas comisionistas más importantes y grandes del país. Además, posee una gran trayectoria prestando servicios en el mercado de valores.

La regulación de la Superintendencia es muy específica al indicar que las entidades vigiladas deben identificar los riesgos asociados a cada uno de los 4 factores de riesgo delimitados en la norma. Para esto debe implementar metodologías de segmentación para cada uno de los cuatro factores de riesgo, los cuales son:

- **Cliente:** Es toda persona natural o jurídica con la cual la entidad establece y mantiene una relación contractual o legal para el suministro de cualquier producto propio de su actividad.

- **Producto:** Son las operaciones legalmente autorizadas que pueden adelantar las entidades mediante la celebración de un contrato.
- **Jurisdicción:** es el punto geográfico relacionado con la operación financiera.
- **Canal de Distribución:** es el medio por el cual se distribuyen los productos o se realizan las operaciones financieras.

Esta segmentación consiste en implementar metodologías estadísticas que permitan fragmentar cada factor de riesgo en una serie de grupos o segmentos homogéneos respecto a una o varias variables. Estos segmentos permitirán aplicar controles específicos según las características y movimientos de los miembros de cada segmento, permitiendo prevenir y detectar riesgos asociados de una forma más efectiva.

Ahora bien, por ser entidades financieras de primer piso, es decir, que prestan sus servicios al público en general, el punto crítico de estas metodologías de segmentación se encuentra en el factor de riesgo clientes, ya que de la correcta segmentación de este factor de riesgo en segmentos homogéneos en su interior y lo suficientemente heterogéneos entre sí, es que se podrían identificar movimientos individuales inusuales y así identificar las mayores amenazas antes de que utilicen el sistema financiero como instrumento de “blanqueamiento de dinero”.

Por otro lado, es interés del presente estudio saber qué tan coherentes pueden ser estas metodologías para identificar movimientos inusuales relacionados con el lavado de activos y la financiación del terrorismo, con la coyuntura económica que atraviesan los diferentes sectores de la economía de los que pueden ser parte los clientes de las entidades financieras. Es decir, que más allá de indagar por metodologías estadísticas de segmentación, se insta por evaluar que tan consistentes son los potenciales segmentos encontrados por estas metodologías, con el contexto económico que atraviesan los diferentes sectores de la economía del país, para así, poder detallar a la luz de las perspectivas de análisis económico, que tan afines pueden ser estas metodologías con las lógicas de los diferentes mercados.

En consecuencia, el estudio pretende identificar por cada uno de los sectores económicos que más presenten alertas durante el periodo de estudio (los 9 primeros meses de del año 2017), cuáles de estos presentan comportamientos positivos, negativo o estables, para así, otorgar un mayor o menor grado de validez a la alerta generada. Verbigracia el estudio podría indicar que las actividades económicas en el sector de manufactura están presentando movimientos financieros atípicos a su perfil de comportamiento; pero a su vez este sector de la economía podría encontrarse en un momento boyante por un aumento en la demanda de este tipo de bienes.

La situación anterior podría sustentar los movimientos atípicos de este sector en su objetivo de incrementar la producción para ajustar la oferta del mercado con la demanda. En este caso la alerta no tendría mayor importancia porque se explicaría en sus propias dinámicas. Pero si por el contrario, este sector de la economía llegase a presentar resultados negativos, la alerta tendría una relevancia más significativa, ya que no existiría una coyuntura económica que sustentare el resultado atípico de sus comportamientos financieros.

Es con base a lo anterior, que se entiende que es importante saber cuáles son los factores macroeconómicos y microeconómicos que pueden influenciar en la definición de los segmentos objeto de la metodología SARLAFT; ya que, a partir de estos factores, se podría analizar qué tan integral es la política pública colombiana en contra de los delitos financieros. Es decir, qué tanto exige el gobierno a las entidades financieras conocer a sus clientes. Ya que un control estadístico que permita identificar comportamientos atípicos a un perfil de comportamiento financiero preestablecido, no es suficiente indicio para identificar potenciales crímenes relacionados con el lavado de activos y la financiación del terrorismo.

Para tener suficiente información es necesario conocer los diferentes contextos económicos que pueden envolver a cada uno de los clientes, puesto que las dinámicas económicas regionales o internacionales, los ciclos de la economía de cada sector y los diferentes factores que influyen en la oferta y la demanda sectorial, pueden explicar la segmentación o sustentar el comportamiento atípico retrospectivo para la singularidad de cada cliente.

Para el efecto, existen diferentes estudios que se han cuestionado sobre el desarrollo de “economías a la sombra” o “economías subterráneas” paralelas a la economía formal de un país, las cuales generan imperfecciones en el mercado por la intromisión de recursos que pueden pervertir el flujo normal de los recursos de los agentes económicos, o representar un detrimento al erario público por la evasión de impuestos, controles y regulaciones. Para Colombia, durante el último cuarto del siglo XX se calculó que el tamaño de estas economías subterráneas representaba entre el 20% y el 50% del PIB, teniendo su mayor proporción en la década de los 90’s estimulada por el desempleo y los diferentes impuestos (Schneider & Hametner , 2007).

La UIAF en su informe “La Dimensión Económica del Lavado de Activos” hace referencia específicamente al efecto negativo que pueden representar recursos de origen ilícito:

“La inyección de capitales de origen ilícito en algunos sectores de la economía, tales como el de la construcción, el sector exportador y el sector bursátil han facilitado la generación de condiciones artificiales de bienestar no sostenibles”. (UIAF, 2014)

Estas condiciones artificiales a las que hace referencia la UIAF, son las que sustentan el por qué es necesario identificar los sectores de la economía que tienen mayor riesgo de padecer estas anomalías en el origen de sus recursos, ya que a medida que sea más factible identificar los focos principales por los que estos recursos se filtran en la economía, le resultaría más fácil al gobierno realizar controles y monitoreos específicos para cada sector.

La perspectiva del análisis económico podría significar un segundo paso en la metodología de segmentación exigida en el SARLAFT, puesto que luego de identificar movimientos inusuales en cada uno de los segmentos, se podría enfocar aún más los controles en los segmentos con comportamientos contrarios a la lógica de la economía de un sector, dejando en un segundo plano a los clientes que puedan llegar a tener movimientos inusuales derivados de cambios en las condiciones del mercado a su favor. Esto se explica en la medida que se pueda conocer la composición de los segmentos, sus características más importantes y sus variables determinantes, ya que con esto se podría estudiar que tan coherentes pueden llegar a ser sus resultados frente a la situación económica que pueden atravesar los diferentes sectores de la economía que figuran como clientes y usuarios del sistema financiero.

1.1 Preguntas

Es por todo lo anterior, que surgen dos cuestiones rectoras que servirán como columna vertebral del presente estudio:

- La primera, surge desde una perspectiva metodológica que pretende dar una respuesta práctica a los requerimientos del SARLAFT y trata de dar respuesta a la siguiente pregunta: ¿cómo se podrían identificar señales o indicios de alerta respecto al riesgo de lavado de activos y financiación del terrorismo sobre las transacciones de clientes de entidades financieras en Colombia, de acuerdo con las metodologías de segmentación exigidas por la Superintendencia Financiera de Colombia en el 2017?
- La segunda será una consecuencia de la primera, la cual, luego de dictaminar las variables más importantes para segmentar y monitorear a los clientes de la entidad financiera, buscará responder una segunda pregunta: ¿cómo se podría integrar el análisis de los sectores de la economía a los que pertenecen los clientes con mayor riesgo LA/FT,

con el Sistema de Administración de Riesgos de Lavado de Activos y Financiación del Terrorismo exigido por las autoridades colombianas?

2. Objetivos

2.1 Objetivo general: Identificar movimientos financieros atípicos relacionados al riesgo de lavado de activos y financiación del terrorismo por parte de los usuarios del sistema financiero colombiano mediante la implementación de un modelo de segmentación de minería de datos y el análisis económico sectorial en una entidad vigilada por la Superintendencia Financiera.

2.2 Objetivos específicos

- Aplicar metodologías de análisis clúster para los clientes de la entidad financiera a estudiar con el fin de encontrar segmentos homogéneos.
- Emplear un modelo predictivo de clasificación de árboles de decisión a los resultados de la segmentación, con el fin de conocer el conjunto de reglas de decisión que definen la composición de cada segmento.
- Definir señales de alerta que permitan identificar datos atípicos para cada segmento encontrado.
- Realizar un análisis económico sectorial para los códigos de la Clasificación Industrial Internacional Uniforme (CIIU) de los clientes con naturaleza jurídica que tengan mayor riesgo LA/FT según la metodología de segmentación, con el fin de identificar si los comportamientos del sector de la economía al cual pertenecen son o no coherentes con el segmento definido.

3. Justificación

Para lograr combatir los crímenes financieros, el gobierno colombiano y las entidades que pertenecen al sector financiero deben seguir estrictos protocolos internacionales para sanear los dineros que circulan dentro de la economía. Desde el punto de vista de las entidades financieras, este camino se encuentra labrado por la Circular Básica Jurídica de la Superintendencia Financiera. Esta circular exige que se implementen metodologías de segmentación para la correcta identificación de señales de alerta sobre riesgos relativos al Lavado de Activos y Financiación del Terrorismo. El presente estudio se justifica de una manera metodológica ya que es una propuesta para la ejecución de un modelo estadístico de segmentación y la elaboración de parámetros de alerta exigido por el regulador colombiano. También se justifica de una manera práctica en la medida que se entienda que puede servir de guía a las entidades financieras de Colombia para evitar riesgos relacionados al LA/FT dentro de sus operaciones.

Además de ayudar a cumplir con la normatividad que les exige la implementación de este tipo de metodologías.

En cuanto al aspecto metodológico se justifica también en la proposición de integrar al modelo estadístico una variable que le otorgue mayor contexto económico a la identificación de movimientos de capitales ilícitos. Este nuevo componente podría ayudar en la integración de las metodologías estadísticas de segmentación con la información puntual de los sectores de la economía que le signifiquen mayor riesgo, permitiendo evaluar los riesgos de lavado de activos y financiación de terrorismo de una manera más integral y focalizada.

4. Marco teórico

4.1 Diccionario Conceptual

- **SARLAFT:** Sistema de Administración de Riesgos de Lavado de Activos y Financiación del terrorismo.
- **UIAF:** Unidad de Información y Análisis Financiero.
- **Minería de datos:** técnicas estadísticas para encontrar información en conjuntos de datos de grandes magnitudes.
- **Análisis clúster:** método que pretende agrupar observaciones según criterios de homogeneidad descifrados en variables de observación.
- **Análisis clúster No jerárquico:** métodos de optimización en los que se indica a priori el número de segmentos. Realizando la distribución de las observaciones con un criterio de optimización del criterio de selección.
- **Criterio de selección:** medida usada para segmentar las observaciones según un conjunto de variables que caracterizan cada una de las observaciones.
- **K medias:** técnica de segmentación que pretende dividir un conjunto de n observaciones en k segmentos homogéneos entre sí y heterogéneos frente a los demás, por medio de una medida de distancia.
- **Distancia euclidiana:** es el insumo principal para el algoritmo k medias, establece un parámetro de medición entre observaciones.
- **Métodos de aprendizaje no supervisado:** son métodos de clasificación útiles para encontrar patrones que permitan caracterizar un conjunto de información.

- **Árboles de decisión:** es un método de aprendizaje no supervisado útil para encontrar las reglas que definen los segmentos. Estas reglas definen las características con las cuales se pueden direccionar con cierto grado de predicción las observaciones estudiadas.

4.2 Marco Teórico y antecedentes:

El contexto expuesto en el planteamiento del problema indica el camino por el cual se dirigirá el presente estudio. Éste se desarrolla por medio de un modelo que sirva para cumplir con la etapa de identificación del SARLAFT por parte de una entidad financiera perteneciente al sector del mercado de valores. Por tener naturaleza relativa al mercado de valores, la entidad financiera objeto de estudio debe garantizar a sus reguladores (AMV, Superintendencia Financiera y Banco de la República) que los recursos que están bajo su administración tienen los controles necesarios para evitar el lavado de activos y financiación del terrorismo.

Al mercado de valores acuden ahorradores, fondos de pensiones, inversionistas extranjeros, inversionistas profesionales, instituciones públicas, entre otros, para invertir sus recursos en instrumentos financieros propios de éste mercado. Estos instrumentos transmiten los recursos de estos oferentes de capital a demandantes de recursos que necesitan capital para cumplir con su objeto social. Estos demandantes representados en empresas acuden al mercado de valores como una alternativa al sistema bancario tradicional, por lo tanto, esperan que los recursos que de allí se pueden conseguir cumplan con las mismas condiciones de sanidad que el sistema bancario ofrece. De allí surge la necesidad de que las entidades financieras del mercado de valores cumplan a cabalidad con sistemas de administración del riesgo LA/FT con respaldo internacional.

El SARLAFT que se ejecuta en Colombia comenzó a exigirse desde 2008 por la Superintendencia Financiera, su composición está diseñada de acuerdo a las recomendaciones realizadas por la institución FAFT (Force Financial Action Task), el cual por medio de un documento oficial llamado “International Standards on Combating Money Laundering and The Financing of Terrorism and Proliferation” establece la metodología por la cual los diferentes gobiernos que pretenden exigir controles sobre el lavado de activos y financiación del terrorismo dentro de su territorio, deben guiarse para diseñar la regulación y el sistema que deben cumplir quienes aspiren a ejecutar sus actividades dentro del país (Financial Action Task Force (GAFI), 2016).

Esta misma institución es la que permite saber cómo los riesgos relativos al LA/FT pueden materializarse, esta definición de riesgos se encuentra expresada en el documento oficial llamado “National Money Laundering and Terrorist Financing Risk Assessment”, de acuerdo a este documento la identificación de riesgos elaborado en un SARLAFT debe tener la capacidad de identificar individuos y/o organizaciones que realicen movimientos financieros que representen una amenaza o un riesgo de “blanqueamiento” o legalización de dineros ilícitos (Financial Action Task Force (GAFI), 2013).

Teniendo claro cuáles son los antecedentes que sustentan la ejecución de este tipo de modelos de identificación de riesgos LA/FT, es necesario mencionar que existen muy pocos estudios que desarrollen una metodología específica de segmentación de factores de riesgo LA/FT. Tales como los estudios realizados sobre las metodologías de minería de datos útiles para el LA/FT elaborado por la UIAF, 2014) y la proposición de técnicas de minería de datos para prevenir LA/FT de Santamaría Ruiz, 2017. Ambos proponen técnicas, pero ninguno las desarrolla. Frente a este tema se encuentran guías estatales como el documento oficial de la UIAF llamado “Técnicas de Minería de Datos para la Detección y Prevención de Lavado de Activos y Financiación del Terrorismo (LA/FT)” que hacen recorridos sobre las diferentes metodologías que puede aplicar las entidades financieras.

Las técnicas de minería de datos más efectivas para la implementación de un SARLAFT son las técnicas descriptivas como lo son la segmentación de observaciones (k-medias) y las técnicas predictivas que intentan establecer cuáles son las características que dicen cómo serán la composición de los segmentos (Santamaría Ruiz, 2017). De este tipo de documentos y de otras lecturas realizadas, es que se logró establecer que las técnicas más efectivas para conseguir segmentos homogéneos en su interior y heterogéneos en su exterior, tal cual lo exige la ley, éstas estarán explicadas en el numeral de la metodología.

Por su parte, la integración de este tipo de modelos con análisis económico por sectores es un campo poco explorado, puesto que por defecto no se suele relacionar la una con la otra fácilmente. Mientras que las metodologías de segmentación procuran determinar patrones de comportamiento por segmentos que evalúen movimientos atípicos dentro de los clientes de una entidad financiera, el análisis económico por sectores indaga por conocer las dinámicas económicas que rodean a un sector de la economía.

Lo anterior, a priori, podría suponer polos extremos para una metodología de investigación, pero si se entiende que las dinámicas de la economía son muchas veces, independientes entre sectores, podría analizarse qué tan probable o coherente es que un cliente cambie su comportamiento financiero. Como por ejemplo el empresario que dobló su producción y por ende sus estados financieros, debido a una sobre demanda en sus bienes y servicios ofrecidos, o el empresario que sin afectar su producción o sus ventas presenta el doble de sus ingresos.

4.3 Planteamiento de la hipótesis.

La hipótesis del estudio consiste en demostrar que es viable aplicar modelos de minería de datos en el cumplimiento de prescripciones normativas frente a la prevención del lavado de activos y financiación del terrorismo en Colombia. Esto supone que la información objeto de este estudio, con todas sus variables y sus especificaciones normativas, tenga una solución clúster por medio de la metodología de k medias, y que a su vez esta solución pueda ser explicada por medio de un árbol de decisión, el cual determinará las características de cada uno de los segmentos, para luego establecer un estadístico para cada uno que permita identificar transacciones atípicas dentro de sus miembros. Además, se tiene la hipótesis de que estos resultados de movimientos inusuales deberían ser incoherentes e inconsistentes con las dinámicas económicas de los sectores de la economía a los que pertenecen sus clientes, ya que una industria con variaciones recesivas en sus indicadores económicos no debería presentar información financiera que muestre lo contrario.

5. Metodología

5.1 Descripción de la información

5.1.1 Variables⁸

Las variables que serán usadas para desarrollar el estudio están taxativamente consagradas en la Circular Básica Jurídica de la Superintendencia Financiera, específicamente en la Parte I, Título IV (la cual compone la regulación colombiana relativa al tema) estas son:

- **Edad:** Por ser una medida cuantitativa que describe algunas características de la población, la edad resulta útil a la hora de construir los clusters, ya que configura una

⁸ Con el objetivo de dar mayor credibilidad a la información usada como insumo para las variables del modelo de segmentación, es imperativo dejar claro que esta información, y en especial la financiera es presentada por todos los tipos de clientes de la entidad estudiada por medio de una serie de reportes y papeles oficiales que hacen constar que esta información es fidedigna y es presentada bajo preceptos legales que garantizan su confiabilidad.

variable fidedigna para comparar comportamientos de sujetos según su madurez económica.

- **Información financiera:** Ingreso Mensual, Egreso mensual, Activos, Pasivos y Patrimonio. Todos reportados y sustentados por el cliente.
- **Frecuencia de transacciones⁹:** esta variable corresponde a la frecuencia anual con la que los clientes realizan transacciones.
- **Valor de transacciones:** esta variable corresponde al volumen anual con el que realizan transacciones los clientes, este volumen se calcula con el promedio ponderado de las transacciones que signifiquen un movimiento de recursos dentro de la entidad financiera con dineros entregados anteriormente.

Estas variables estarán disponibles para cada una de las observaciones, las cuales reposan en una base de datos de clientes aportadas por una entidad financiera. Cómo su tratamiento es reservado y cualquier tipo de manipulación de este tipo de información está protegida por diferentes derechos de confidencialidad, la información será proporcionada única y exclusivamente para la realización de la metodología con fines académicos. En ningún momento se podrá conocer detalle de cada observación asociado a un nombre o número de identificación, lo que garantizará la confidencialidad de la información y protección de los clientes. A su vez que no afectarán al estudio en la medida que solo necesita la información de las variables sin necesidad de identificar a quien pertenece la información.

5.1.2 Observaciones

La metodología de segmentación será aplicada sobre 7747 observaciones, las cuales están representados en 6319 clientes con personería natural y 1428 personas jurídicas. Si bien el procedimiento se desarrollará de manera paralela para cada tipo de personas, es necesario aclarar que para las personas jurídicas se cuenta con mayor precisión en la información disponible, ya que, para ser clientes de la entidad financiera debieron proporcionar con un nivel de detalle de 4 dígitos el Código de Clasificación Internacional

⁹ Para el procedimiento inicial que comprende los años 2015 y 2016, esta variable se llama Frecuencia _T_a para el año 2015 y Frecuencia_T para el año 2016.

Industrial (CIIU)¹⁰. Las personas naturales solo proporcionan su ocupación o la actividad económica que desempeñan (asalariado, pensionado, independiente, etc.).

5.1.3 Software implementado

Para poder aplicar las técnicas de minería de datos usadas en el estudio, se optó por hacer uso de la herramienta informática llamada R Project. Esta herramienta tiene un lenguaje propio, el cual funciona de manera colaborativa y en el que se puede hacer uso de funciones previamente programadas por aportes de otros usuarios alrededor del mundo. Esta herramienta resultó muy útil para hacer uso de las funciones necesarias para el modelo, además permitió que el mismo quedara plasmado en un código que facilita su futura replicación.

5.2 Descripción de la metodología seleccionada

En cuanto a la metodología, para poder elaborar un modelo estadístico de segmentación por medio de técnicas de minería de datos, se debe acudir a las técnicas de clusterización y árboles de decisión. Puesto que el objeto de las metodologías de segmentación es agrupar objetos similares, se necesita alguna medida para evaluar las diferencias y similitudes entre objetos. El concepto de similaridad es fundamental en el Análisis Cluster. La similaridad o similitud es una medida de correspondencia o semejanza entre los objetos que van a ser agrupados, la estrategia más común consiste en medir la equivalencia en términos de la distancia entre los pares de objetos. Esta distancia es conocida como distancia euclidiana, la cual se obtiene de medir el espacio que hay entre un par de objetos según los valores de sus variables y una métrica específica. Lo anterior se hace por medio de una aplicación del teorema de Pitágoras (Ghanshyam, 1991). La distancia euclidiana se define en la siguiente ecuación:

$$d(p, q) = \sqrt{(q_1 - P_1)^2 + (q_2 - P_2)^2 + \dots + (q_n - P_n)^2}$$

Donde p y q son los pares de objetos observados y son comparados según su ubicación en un espacio multidimensional. (Jhonny R. Demey, 2011)

Ahora bien, la distancia euclidiana puede otorgar una medición entre observaciones con variables cuantitativas, pero para el caso del modelo de segmentación del estudio es necesario tratar con variables cualitativas como la ocupación o sector de la economía de los clientes. Para solucionar esta situación la medida de distancia seleccionada es la Distancia de Gower, la cual

¹⁰ Por tratarse de códigos CIIU de máximo 4 dígitos, además de ser obtenidos a partir de la información financiera proporcionada por los clientes jurídicos que sustentan su situación según informes oficiales reportados por ellos a los órganos reguladores de Colombia, los cuales como mínimo están actualizados en al menos 2 años antes del 2016; es pertinente aclarar que la versión de la clasificación industrial usada para este estudio corresponde a la revisión 4 adaptada para Colombia. (Departamento Administrativo Nacional de Estadística (DANE), 2012).

tiene la virtud de medir distancias entre objetos que posean variables tanto cuantitativas como cualitativas. Esta distancia se encuentra definida como una posterior modificación a la distancia euclidiana, la cual se encuentra definida de la siguiente manera por la ecuación $d_{ij}^2 = 1 - S_{ij}$, i y j las observaciones. S_{ij} sería el coeficiente de Gower (Gower, 1971) definido de la siguiente manera:

$$S_{ij} = \frac{\sum_{h=1}^{P_1} \left(1 - \frac{|X_{ih} - X_{jh}|}{G_h} \right) + a + \alpha}{P_1 + (P_2 - d) + P_3}$$

Donde P_1 es la cantidad de variables cuantitativas continuas, P_2 es la cantidad de variables binarias, P_3 es la cantidad de variables cualitativas, a y d son las coincidencias (1, 1) en las variables binarias, α son las coincidencias en las variables cualitativas y G_h el rango que contiene la n-ésima variable observada. Esta distancia nos permite comparar objetos con variables mixtas.

Ahora bien, la distancia de Gower otorga una definición de distancia a todos los individuos observados, para que después en forma matricial, se determine la distancia de cada individuo en relación con todos los demás individuos estudiados. Dicha matriz determina cual será el grado de similitud entre individuos y permite que la segmentación de grupos cumpla con el requisito de que sean homogéneos en su interior y heterogéneos en su exterior (Kaufman & Rousseeuw, 2005). Esta matriz de distancias se encuentra definida de la siguiente manera:

$$\begin{matrix} & \begin{matrix} p \text{ variables} \end{matrix} \\ \begin{matrix} n \text{ objects} \end{matrix} & \begin{bmatrix} x_{11} & \cdots & x_{1f} & \cdots & x_{1p} \\ \vdots & & \vdots & & \vdots \\ x_{i1} & \cdots & x_{if} & \cdots & x_{ip} \\ \vdots & & \vdots & & \vdots \\ x_{n1} & \cdots & x_{nf} & \cdots & x_{np} \end{bmatrix} \end{matrix}$$

Fuente: (Kaufman & Rousseeuw, 2005)

Esta matriz de distancias es el insumo principal del siguiente paso de la metodología, para el estudio fue seleccionada la técnica “K-means” propuesta a finales de los 60’s por James MacQueen, la cual consistía en fragmentar un conjunto de observaciones con tamaño N $X_1 X_2, \dots X_N$, en k conjuntos que minimicen la suma de cuadrados de todos los objetos (MacQueen, 1967). A partir de la definición de un número de centroides seleccionados al azar, el algoritmo k medias calcula la distancia entre todos los objetos observados y define los segmentos según la ubicación de esos centroides. Este método de clusterización se conoce como un método jerárquico, ya que se debe ingresar cual es el número inicial de k con el que se quiere hacer la segmentación.

Sobre esta teoría de clusterización, Leonard Kaufman y Peter Rousseeuw propusieron un algoritmo mejorado llamado “*Partitioning Around Medoids (PAM)*”. Este algoritmo haría lo mismo que el de k medias, pero encontrando las observaciones de una manera más robusta, encontrando los k segmentos más representativos de los datos y asignando cada observación al objeto representativo que más se parezca (Kaufman & Rousseeuw, 2005). Los objetos con distancias reducidas entre ellos son más parecidos entre sí, que aquellos que tienen distancias mayores, los más similares son los que se agrupan para conformar clústers. La parte práctica de este algoritmo fue presentada en el 2006 por Graeme Richards, Beatriz de la Iglesia y Victor Rayward-Smith en la comparación de metodologías de clusterización jerárquicos y no jerárquicos de división, en esta publicación mostraron como se podía aplicar el método PAM encontrado la cantidad óptima de segmentos (Richards, De la Iglesia, & Rayward-Smith, 2006).

Es necesario mencionar que el algoritmo de las k medias, ejecutado desde a metodología PAM es un método de análisis clúster no jerárquico, lo que quiere decir que a priori no se conoce el número óptimo de clústers por el que estará compuesto la información. Este número sólo se obtiene en métodos jerárquicos por medio de la creación de dendogramas que lo muestren gráficamente. Ahora bien, cómo este tipo de métodos no fueron concebidos dentro del estudio por el gran tamaño de la población y por la necesidad explícita en la norma de obtener segmentos homogéneos en su interior y heterogéneos en su exterior, es imperativo encontrar una forma de indicar al método no jerárquico seleccionado, cuál será el número óptimo de segmentos.

Para llegar a una solución óptima, k- medias se realiza por medio del algoritmo PAM varias veces con diferentes “siluetas”¹¹ que representan el posible número de clústers que compondría el conjunto de información. Estas siluetas se miden a través del cálculo de distancias entre los individuos estudiados y centroídes escogidos aleatoriamente dentro de la población, los cuales se definirán entre 2 y 35 posibles clústers para clientes personas jurídicas y entre 2 y 20 para personas naturales¹², de las cuales, será seleccionada la que mejor explique el comportamiento de los datos. Esta selección se hace por medio de un gráfico que muestra el resultado de cada una de las siluetas según el número de clústers y un índice de “efectividad” para cada silueta otorgado por el mismo algoritmo. Los límites de los posibles rangos de clústers se definen bajo

¹¹ Las siluetas son la ejecución del algoritmo PAM por cada número de clústers dentro de un rango preestablecido. Estas siluetas tienen la bondad de explicar por medio de un índice, que cantidad de variabilidad de la data es explicada por cada número de clúster.

¹² Se definen esta cantidad máxima de posibles clústers dado a la variabilidad entre estos diferentes tipos de clientes, son mucho más diferentes los jurídicos que los naturales.

la premisa de ensayo y error sobre las diferentes ejecuciones del modelo, partiendo que el mínimo debe ser 2 puesto que la información debe ser particionada por lo menos una vez, y el superior en qué en ninguno de los casos estudiados una silueta por encima de 35 para jurídicos y 20 para naturales supera una importancia de más del 50% del índice de efectividad.¹³

Luego de obtener el número de clústers que mejor describen la información, se usará el algoritmo llamado T-SNE, el cual resulta útil en procesos de reducción de dimensiones de grandes conjuntos de datos. Para el caso de estudio resulta útil en la medida que permite graficar en un plano cartesiano la ubicación que tendría cada una de las observaciones según la información de cada una de sus variables. El gráfico generado contendrá una descripción visual del comportamiento todas las observaciones y su ubicación en el número de clústers encontrado en el paso anterior (Van der Maaten & Hinton, 2008).

Ahora bien, la definición de estos clústers para cada observación se hace en razón de crear una variable dependiente, con la que se pueda aplicar un algoritmo de clasificación para conocer la composición de los segmentos. Para esto se presentan dos métodos de aprendizaje no supervisado conocidos como “Decisión Tree Learning”, en específico los algoritmos de clasificación “rpart” (Breiman, Friedman, Olshen, & Sontag, 1984) y “C5.0” (R., 1993). Estos algoritmos funcionan como “machine learning”, en los que se hacen uso de la técnica de árboles de decisión en un modelo de clasificación predictivo para definir las reglas de los segmentos que representan la distribución óptima de los individuos.

En esta etapa se obtiene una lista con las reglas que definen los segmentos. Estos segmentos serán el producto final de la metodología de segmentación, y representarán los segmentos óptimos en los que mejor se distribuyen las observaciones. Por último, para poder definir los valores extremos para el caso de estudio, se replicará la metodología desarrollada por “M. Hubert” y “E. Vamderieveren”, de la Universidad de Leuven (Bélgica) y la Universidad de Antwerp (Bélgica), respectivamente, quienes en su investigación titulada “An Adjusted Boxplot for Skewed Distributions” desarrollan estudios sobre cómo definir un rango basado en el concepto de cuartiles que identifique los límites inferiores y superiores en los cuales se concentraría los valores de una muestra (Hubert & Vandervieren, 2007). Por tratarse de un

¹³ Estas pruebas fueron realizadas con un tope máximo de 100 posibles clústers, partiendo desde la premisa de entender que entre más segmentos llegase a tener el conjunto de datos más útil sería su segmentación, pues el número k de segmentos y n de observaciones sería más cercano. Reduciendo el margen de acción del estudio a un análisis individual por cada observación.

análisis con variables ordinales como el dinero, no será usado el concepto de límite inferior, ya que se el piso del rango de monitoreo siempre será \$0.

Los estadísticos propuestos en la mencionada investigación se basan en el concepto estadístico de “Medcouple” (Brys, Hubert, & Struyf, 2004). Este concepto fue introducido en 2004 y es de gran utilidad como método de medida de asimetría, ahora bien, este estadístico proporciona un resultado positivo o negativo, según sea el caso, los autores mencionados proponen los siguientes estadísticos propiciados por su investigación:

Cuando se trate de valores atípicos leves y el $MC \geq 0$:

$$Q_1 - 1.5 e^{-4MC} IQR ; Q_3 + 1.5 e^{3MC} IQR$$

Cuando se trate de valores atípicos leves y el $MC < 0$:

$$Q_1 - 1.5 e^{-3MC} IQR ; Q_3 + 1.5 e^{4MC} IQR$$

Cuando se trate de valores atípicos extremos y el $MC \geq 0$:

$$Q_1 - 3 e^{-4MC} IQR ; Q_3 + 3 e^{3MC} IQR$$

Cuando se trate de valores atípicos extremos y el $MC < 0$:

$$Q_1 - 3 e^{-3MC} IQR ; Q_3 + 3 e^{4MC} IQR$$

Relativo a las fórmulas definidas para calcular los estadísticos que definirán el límite de transacciones realizadas por los clientes pertenecientes a un segmento, se define que las variables serán, para cada segmento en particular, el primer cuartil (Q1), el cual será el inicio del rango operado por el MC= medcouple y el IQR= rango intercuartílico. Consecuentemente, el tercer cuartil (Q3) será tratado de manera semejante según el caso. Es necesario aclarar que estos estadísticos abarcan distribuciones con valores tanto positivos como negativos. Esta cualidad no será necesaria en el presente estudio, puesto que, por tratarse de una variable nominal, como lo es el dinero, solo es de interés conocer cuál es el límite superior del rango de transacciones. Es por la creación de estos rangos, que se puede definir el espectro “normal” en el que los miembros de cada segmento se comportan, para sí definir posibles alertas.

Al conocer los resultados por segmento, se procederá a elegir las alertas de los clientes personas jurídicas¹⁴ para aplicar pruebas de comportamiento sectorial de sus integrantes. Lo anterior, es

¹⁴ Solo se elige a las personas jurídicas ya que poseen la variable actividad económica CIU.

posible gracias al análisis individual realizado por los códigos CIIU con mayor recurrencia de alertas determinadas según la metodología de segmentación. A estos CIIU se les hará un estudio de comportamiento histórico durante el año 2016 y 2017 para conocer como ha sido su comportamiento dentro de la economía del país, su participación en la misma, su comportamiento cíclico o contra cíclico y sus principales variables macroeconómicas, para luego determinar qué tan riesgosos son éstos a la lógica de los resultados de la segmentación y que tan correlacionadas pueden estar una metodología frente a la otra.

Es necesario mencionar, que, dadas las cualidades de la base de datos disponible, el estudio procurará alinear su alcance con la nomenclatura de los códigos CIIU a 4 dígitos, ya que los registros de la base proporcionan este nivel de detalle y esto podría ayudar a conocer con mayor precisión los diferentes sectores y subsectores economía los cuales pertenecen los clientes de la entidad financiera bajo estudio.

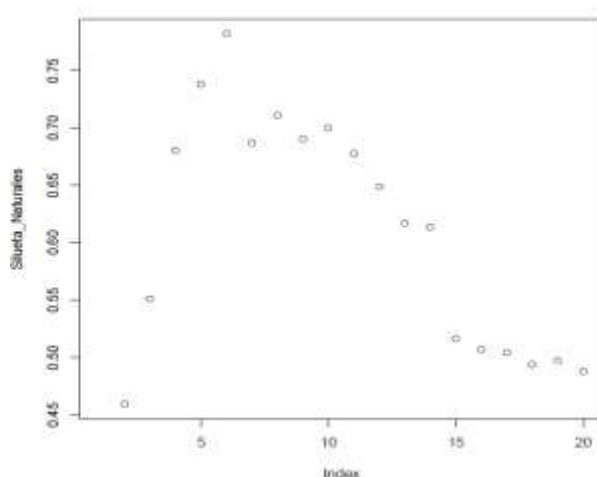
6. Resultados de la metodología de segmentación

Como bien se dijo anteriormente, la metodología descrita es aplicada de manera independiente tanto para clientes naturales como para clientes jurídicos.

6.1 Personas naturales

En el gráfico 2 se encuentra la ejecución del algoritmo PAM entre 2 y 20 posibles siluetas. Se puede ver que la información se explica de mejor manera en pocos clústers, teniendo su punto más alto en 6. Según este gráfico, al segmentar la información en 6 clústers se obtiene un índice de explicación de aproximadamente 0.79. Si se toma en cuenta que 1.0 es el máximo valor que puede tomar este eje del gráfico, queda muy claro que para ejecutar el PAM, 6 clústers será el número óptimo de segmentos que permite clasificar a las personas naturales.

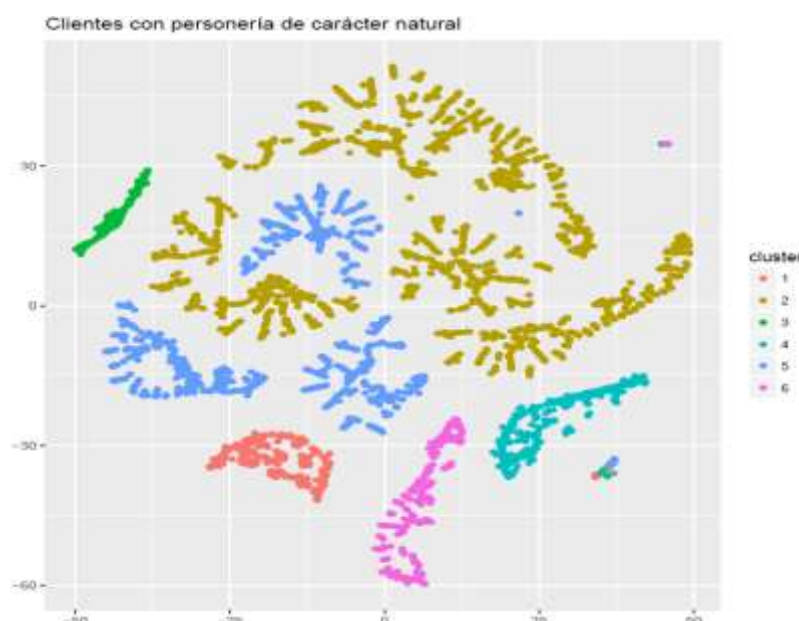
Gráfico 2. Silueta PAM para personas naturales.



Fuente: información de clientes proporcionada por la entidad financiera, 2017. Cálculos propios.

Ahora bien, conociendo el número de clústers que fragmentarán la información, se ejecuta el TSNE¹⁵ para personas naturales, el cual se encuentra descrito en el gráfico 3. Este gráfico muestra que los clústers que componen la información se encuentran bien concentrados alrededor de sus centroides y por lo tanto tienen un alto nivel de “pureza” entre ellos, es decir, arrojan un resultado coherente con el índice de explicación que sugirió escoger la silueta con 6 clústers y se puede diferenciar claramente uno del otro, cumpliendo la premisa rectora de estudio que es encontrar segmentos homogéneos en su interior y lo suficientemente heterogéneos en su exterior.

Gráfico 3. T-SNE Naturales (t-distributed stochastic neighbor embedding).

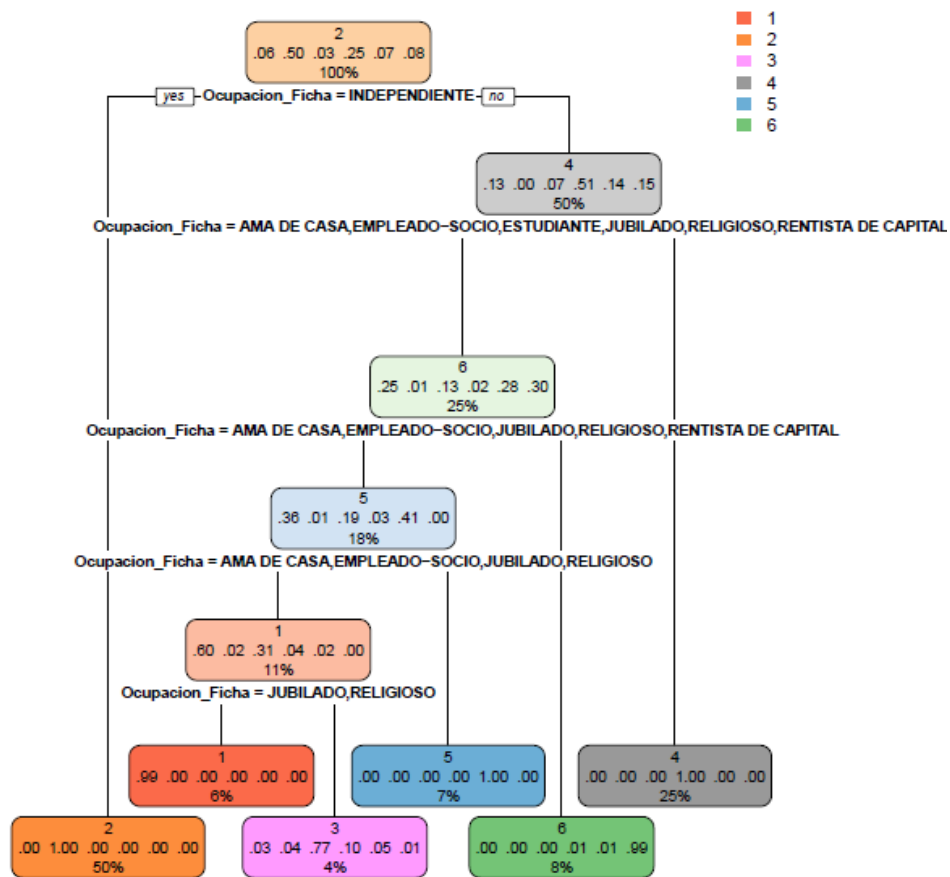


Fuente: información de clientes proporcionada por la entidad financiera, 2017. Cálculos propios.

Los clústers que componen la información, serán el insumo necesario para aplicar los métodos de aprendizaje no supervisado de árboles de clasificación. Lo anterior, en la medida que se entienda que cada cliente con personería natural tendrá una nueva variable que irá de 1 a 6 según el clúster que le haya sido asignado. Las reglas que definen los segmentos encontrados por el algoritmo de clasificación “C5.0” se encuentran descritos en el anexo 1, en estos se puede observar que las principales variables a la hora de segmentar personas naturales son la edad y la ocupación.

¹⁵ Esta técnica es utilizada para reducir data de grandes dimensiones en un plano cartesiano. Su función es únicamente de visualización, por lo que las unidades de sus ejes son arbitrarias según la información de las diferentes observaciones, por lo tanto, su interpretación no se da en términos de métricas de distancia sino en concentración o desconcentración.

Gráfico 4. RPART personas naturales.



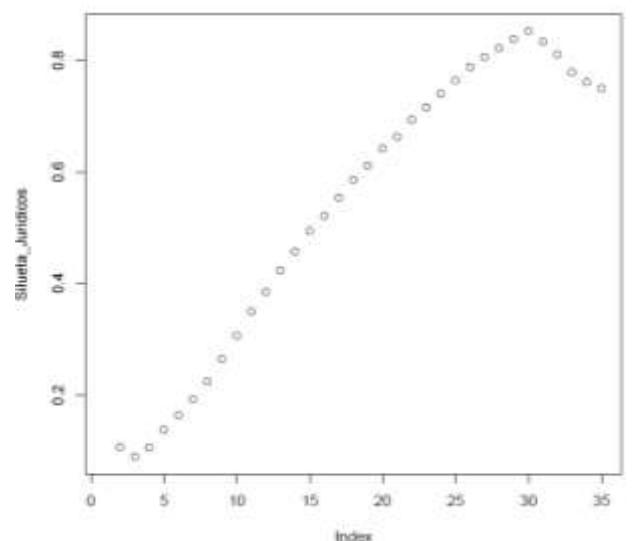
Fuente: información de clientes proporcionada por la entidad financiera, 2017. Cálculos propios.

En cuanto a la generación de alertas para los segmentos naturales, se utiliza el histórico de transacciones de los clientes para el año 2016, con esta información se procede a encontrar un límite superior mensual de volumen transado en un mes para cada segmento.

6.2 Personas jurídicas

Las siluetas del algoritmo PAM para las personas jurídicas tienen un comportamiento muy diferente a las siluetas de las personas naturales. En el gráfico 4 se puede observar que el aumento de la cantidad de clústers es coherente con un mejor índice de efectividad mayor, dicha tendencia se estabiliza en 30 clúster, número en el cual el índice de efectividad adquiere su valor óptimo en aproximadamente 0.85, y en el cual, se comienza a observar una tendencia descendente. Este comportamiento de campana puede estar explicado en la variabilidad por la que está compuesta la información de las personas jurídicas, ya que, por existir mayor distanciamiento entre las variables de la segmentación de las observaciones, es necesario tener un mayor número de clústers que representen la información.

Gráfico 5. Silueta PAM para personas naturales.

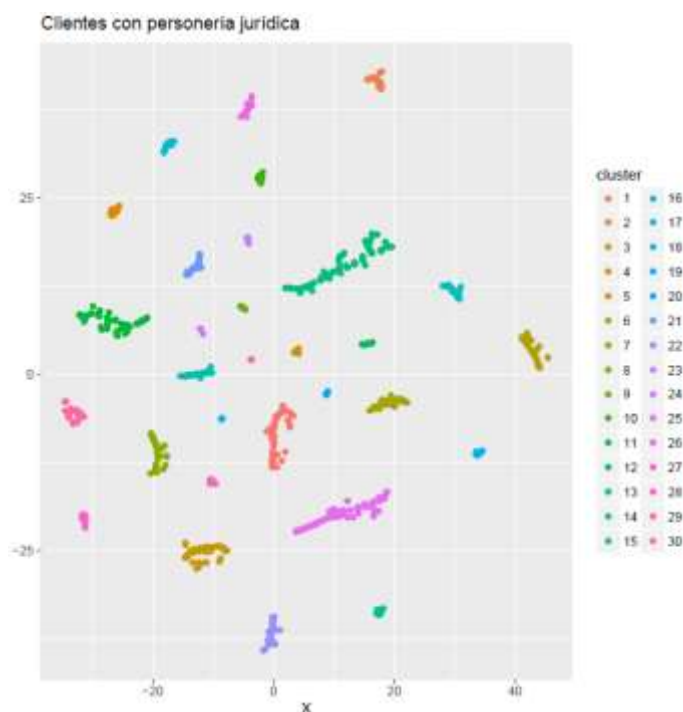


Fuente: información de clientes proporcionada por la entidad financiera, 2017. Cálculos propios.

A pesar de ser un grupo de observaciones más pequeño que las personas naturales, la necesidad de tener un mayor número de clústers que expliquen la información, tiene una posible sustentación en qué su información tiene un menor grado de similitud explicado en los diferentes tamaños, negocios y actividades en los que pueden estar enfocadas las diferentes empresas que son clientes de la entidad financiera. Una muestra gráfica de la anterior suposición se encuentra en el gráfico 5. En este gráfico se puede apreciar que las observaciones se encuentran mucho más distanciadas unas de otras, y por lo tanto debe existir un mayor número de clústers que representen las observaciones. De esta gráfica también se puede concluir que los segmentos de los clientes jurídicos están compuestos por menos observaciones, pero estas componen de manera muy coherente cada uno de los segmentos, esto debido a que la integralidad del color de cada clúster se mantiene “puro” y demuestra que sus observaciones son lo suficientemente homogéneas entre sí.

Ahora bien, por la cantidad de segmentos, el algoritmo “rpart” no puede otorgar una descripción gráfica tan coherente de las reglas que definen los segmentos como lo hace con las personas naturales. A pesar de esto “C5.0” sí puede proporcionar las reglas que definen las características de cada uno de los segmentos encontrados. Para este tipo de clientes el algoritmo definió que la única variable que será determinante para establecer las reglas de los segmentos sería la actividad económica CIIU.

Gráfico 6. T-SNE jurídicos (t-distributed stochastic neighbor embedding).



Fuente: información de clientes proporcionada por la entidad financiera, 2017. Cálculos propios.

El resultado al detalle de cada uno de los segmentos se encuentra en el anexo 2, en esta tabla se puede apreciar que para las personas jurídicas los segmentos tienen un número de integrantes más consistente, el cual, está sustentado en las actividades económicas que reportan los clientes que lo componen, y que, a su vez, permiten la generación de alertas más enfocadas para cada segmento encontrado. También es relevante mencionar que varios segmentos están compuestos únicamente por un código CIIU, esto quiere decir, que las personas jurídicas que desarrollan actividades en este sector de la economía presentan una información muy particular que no se asemeja a sus otros pares.

6.3 Alertas personas jurídicas

Ahora bien, para cumplir con el objetivo final de la presente investigación, el cual consiste en comparar los resultados de las alertas generadas con la coyuntura económica que atraviesan los diferentes sectores del país. Se propuso estudiar solamente las alertas generadas únicamente por las personas jurídicas, puesto que su definición está dada en códigos CIIU y esto permitirá analizar por sectores de la economía que tan coherentes son alertas, o quizá vislumbrar que tan recurrentes son cierto tipos de actividades en la presentación de comportamientos financieros atípicos.

Para ver cuáles son los códigos CIIU que generaron alertas según el modelo, se ejecutó la segmentación con la información presentada para el 2017 y se monitorearon las transacciones realizadas entre enero y septiembre de 2017. De lo anterior, se obtuvo los siguientes resultados:

Tabla 1. ALERTAS PERSONAS JURÍDICAS GENERADAS EN 2017

ALERTAS PERSONAS JURÍDICAS GENERADAS EN 2017									
MES	ENERO	FEBRERO	MARZO	ABRIL	MAYO	JUNIO	JULIO	AGOSTO	SEPTIEMBRE
N° ALERTAS	11	26	18	14	16	17	17	3	14
% TOTAL	8,1%	19,1%	13,2%	10,3%	11,8%	12,5%	12,5%	2,2%	10,3%

Los resultados de las alertas definidas por el modelo de segmentación arrojaron que los códigos CIU que generan mayor número de alertas durante el periodo de estudio son los siguientes:

Tabla 2. Códigos CIU con mayor número de alertas generadas

Clasificación Internacional Industrial Uniforme (CIU)	Alertas
7010- actividades de administración empresarial	39
6810 - actividades inmobiliarias realizadas con bienes propios o arrendados	27
9499 - actividades de otras asociaciones N.C.P.	19
6613 - otras actividades relacionadas con el mercado de valores	16
3091 - fabricación de motocicletas	9
4511- comercio de vehículos automotores nuevos	9
8430 - actividades de planes de seguridad social de afiliación obligatoria	9
5111- transporte aéreo nacional de pasajeros	8

7. Análisis económico sectorial de las alertas generadas

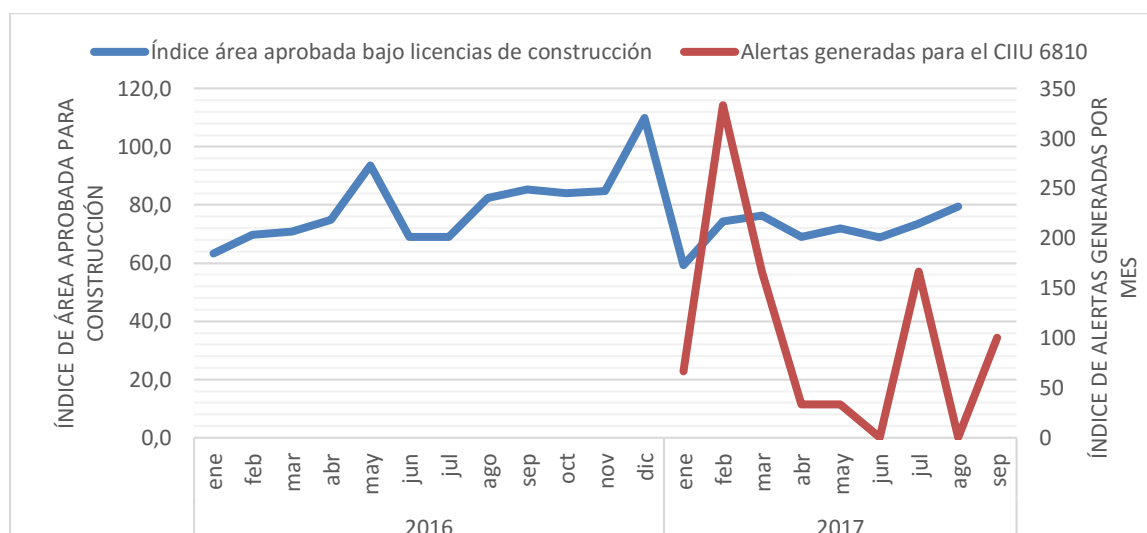
Ahora bien, una vez conocidos los códigos CIU que generan la mayor cantidad de alertas, se puede deducir que éstos son los que representan un mayor riesgo frente al LA/FT para la entidad financiera estudiada. Para poder analizar qué tan coherentes pueden llegar a ser estas alertas con la realidad económica que atraviesa cada uno de los sectores de la economía a los que pertenecen los códigos CIU señalados, y así definir la eficacia de las reglas que los definieron; se comparó el número de alertas generadas mensualmente por cada uno de los códigos CIU con el comportamiento histórico mensual de cada uno de los sectores de la economía a los que pertenecen. Para comparar los sectores de la economía, se usaron series históricas que reflejan el comportamiento que ha tenido cada sector desde enero de 2016 hasta el último mes disponible de cada uno. Estas series se componen por volúmenes de producción, volúmenes de ventas, índices, entre otros.

Para tener un método cuantitativo para comparar la similitud entre las fluctuaciones del comportamiento de la economía y los índices de generación de alertas, se comparó el comportamiento normalizado de cada una de estas series de tiempo, únicamente durante los 9 meses

del periodo de estudio, por medio de un coeficiente de correlación. Estos resultados pueden ser vistos en detalle en el anexo 3.

Ahora bien, los resultados obtenidos para la comparación de cada uno de los índices de generación de alertas, comprada con el comportamiento de su sector, fueron los siguientes:

Gráfico 7. Resultados código 6810 - actividades inmobiliarias realizadas con bienes propios o arrendados.



Fuente: (DANE, 2017). Cálculos propios.

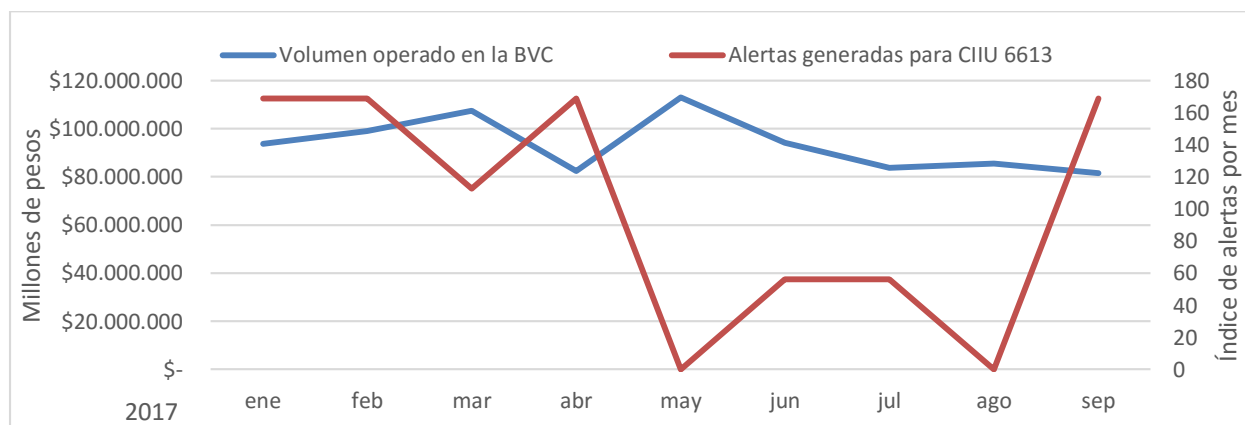
De todos los códigos que arrojaron alertas durante el periodo de estudio, el 6810 es el más volátil de todos. A pesar de esto, el comportamiento histórico de la generación de alertas para este sector es coherente con el comportamiento del mercado inmobiliario. Para realizar esta comparación, fue necesario usar el índice de áreas con licencia para construcción calculado por el DANE. En el gráfico 6 se puede observar cómo para febrero de 2017, mes en el que se generaron 10 alertas provenientes de clientes con este código CIU; aumentó el número de metros cuadrados licenciados para construir. Además de este, los otros movimientos entre ambos índices son consecuentes en sus movimientos. Este comportamiento podría estar explicado en el coeficiente de correlación entre ambas series por valor de 0.22. Lo anterior, en la medida que este coeficiente refleja que ambas series tienen una correlación positiva entre ellas, que pesar de no ser tan alta respaldan el comportamiento antes explicado.

De todo esto, se podría indicar que las alertas generadas por los clientes pertenecientes a este sector de la economía, tendrían un menor grado de relevancia en el propósito de detectar riesgos relativos al LA/FT. Esto en la medida que sus movimientos financieros se explican de algún modo en la coyuntura económica que vive este sector.

En el gráfico 7 se puede observar las alertas generadas por el código CIIU 6613, el cual se dedica a realizar actividades en el mercado de valores. En esta gráfica el comportamiento del índice es contrario al volumen total transado en la bolsa. Para efectos del estudio, el resultado de la comparación entre estas dos series podría sugerir que entre el total del volumen transado en la BVC y las alertas generadas para el CIIU 6613 de los clientes de la entidad financiera estudiada, existe una correlación negativa por un valor de -0.22; esto indica que para cada periodo de tiempo ambas series presentan comportamientos contrarios.

Lo anterior sería muy alarmante en un principio, ya que una correlación negativa entre estas dos series sería una alerta inmediata, pero si se entiende que en este mercado puede existir un rezago de un periodo en el índice de las alertas debido a la naturaleza del mismo, se podría deducir que entre ambas puede existir una relación consecuente, pero con rezago de un periodo. Este comportamiento podría sugerir que el valor del volumen transado en la bolsa antecede a las alertas por un periodo. Esto es, si incrementa el volumen transado para este código en un periodo t , las alertas tendería a aumentar en el periodo $t + 1$.

Gráfico 8. Resultados código 6613 - otras actividades relacionadas con el mercado de valores.

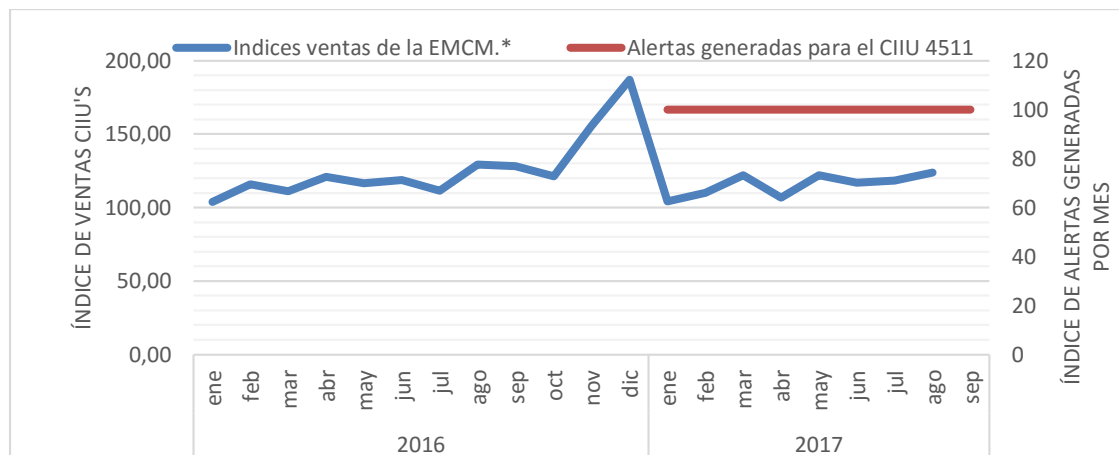


Fuente: (Bolsa de Valores de Colombia, 2017). Cálculos propios.

Para el código 4511, el gráfico 8 muestra los resultados del índice de alertas generadas para este código CIIU comparado con el índice de ventas de vehículos nuevos de la Encuesta Mensual de Comercio al por Menor y Comercio de Vehículos (EMCM) calculado por el DANE. De estos resultados se puede inferir que, a pesar de las fluctuaciones de las dinámicas propias de este sector, los clientes jurídicos de la entidad que están generando alertas, están presentando comportamientos financieros atípicos a su perfil de manera independiente a las alzas y bajas que pueda presentar el mercado. Lo anterior puede indicar que los clientes jurídicos con este código CIIU y que generaron alertas para el periodo de estudio, lo hicieron de manera paralela al comportamiento del sector del

comercio de vehículos automotores. Por lo tanto, el sector de la economía al que pertenecen estos clientes no explicaría el comportamiento financiero atípico de los mismos, elevando a un mayor grado de relevancia las alertas de este sector.

Gráfico 9. Resultados código 4511 comercio de vehículos automotores nuevos.



*Códigos CIIU 4511-4541-4542.

Fuente: (DANE, 2017). Cálculos propios.

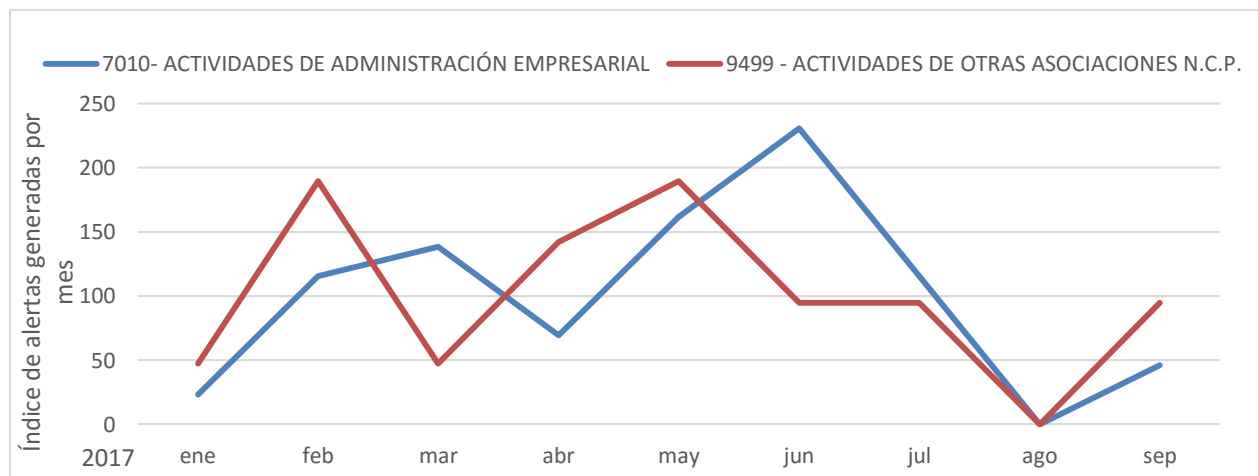
En concordancia con la anterior, para los códigos 3091 - fabricación de motocicletas, 8430 - actividades de planes de seguridad social de afiliación obligatoria y 5111- transporte aéreo nacional de pasajeros, los comportamientos de sus alertas versus el sector de la economía al que pertenecen se comporta de manera equivalente al código 4511. Es decir, la generación de alertas por estos códigos CIIU fueron consistentes durante el periodo de estudio. No variaron según los movimientos económicos de sus sectores respectivos. Para mayor detalle: gráfico 10, 11 y 12 del apéndice.

En forma de ejemplo, se podría aclarar que, para el mercado del transporte aéreo de pasajeros nacionales, las alertas generadas son constante a lo largo del periodo de estudio. Esto es, para los 9 meses estudiados, el código CIIU 5111 generó una alerta para cada uno de estos meses con excepción en el mes de agosto. Esta consistencia en la generación de alertas, está respaldado en la homogeneidad del total de pasajeros transportados durante todos los meses. En el gráfico 11 se puede observar que, desde principios de 2016, el total de pasajeros transportados por vía aérea a nivel nacional se concentra en la media en cerca de dos millones de pasajeros mensuales. El comportamiento de este mercado, comparado con el índice de alertas generadas para este código CIIU podría indicar que las alertas de este sector podrían estar proviniendo de clientes que sostienen movimientos financieros paralelos e independientes a la coyuntura de este sector, y por lo tanto, restando relevancia metodológica a la coyuntura económica de las alertas generadas para este sector.

Por último, para los códigos 7010 y 9499 resultó imposible representar el sector de la economía al que pertenecen por la amplitud de actividades que abarcan las operaciones de los mismos. Para estos códigos, en el gráfico 9, se muestra el índice de alertas generadas mensualmente por cada uno. De este gráfico se puede concluir que la generación de alertas sobre estos códigos es muy volátil y obedece a cierta estacionalidad sincrónica entre ambos. Lo anterior sucede, en la medida que se entienda que en el primer bimestre del año ambos presentaron pocas alertas, para luego tener grandes picos a mitad de año y coincidir con el resto de CIIU's con una mínima presencia de alertas en agosto.

Adicional a la generación activa de alertas provenientes de estos códigos, se suma el riesgo de la naturaleza de las diferentes actividades que se pueden desarrollar desde estos. En primer lugar, el código 9499 sugiere un nivel de riesgo alto para los clientes que tengan este código y presenten alertas. Lo anterior, en la medida que se entienda que este código se adjudica únicamente sobre las personas jurídicas que no logren acomodar sus operaciones y actividades con ningún otro código CIIU, dicha “indecisión” o inconsistencia sobre las actividades económicas desarrolladas por un cliente representan un fuerte indicio, que, adicionado a las alertas generadas por cada uno, podrían indicar que estos clientes son propensos a riesgos LA/FT.

Gráfico 10. Alertas generadas sin sector de la economía específico.



Ahora bien, en cuanto al código 7010 se puede inferir que el riesgo LA/FT de los clientes que desarrollan actividades dentro de los límites de este CIIU están relacionados proporcionalmente con el tamaño de sus movimientos dentro de la entidad financiera. Estos movimientos se traducen en el tamaño que tiene cada una de las empresas que generan este tipo de alertas. Durante el periodo de estudio, 17 clientes con este código CIIU generaron un total de 39 alertas. Al no poder comparar estas alertas con el movimiento de la economía, en la tabla 3 se puede observar cuales son los

tamaños de las empresas que generan alertas según la clasificación prescrita por el gobierno colombiano en la Ley 905 de 2004.

Tabla 3. Cantidad de personas jurídicas con código CIU 7010, que generaron alertas.

Tamaño de empresa	N° empresas que generan alertas
Micro	1
pequeña	4
mediana	7
Grande	5

De esta tabla se puede concluir que las alertas generadas para este código CIU tienen un alto nivel de riesgo LA/FT, esto en la medida que se entienda que la clasificación que prescribe la ley se realiza en función del total de activos que posee cada empresa. Estos activos se traducen en un mayor nivel de recursos con posibilidad de ser introducidos al sistema financiero.

8. Conclusiones

Los controles exigidos por el regulador colombiano en materia del SARLAFT tienden a estar más coordinados y actualizados con los avances académicos e informáticos sobre el manejo masivo de información de clientes por medio de la minería de datos. Las metodologías por las cuales se ejecutan estos controles para prevenir e identificar delitos financieros, cada vez adquieren un mayor grado rigurosidad estadística. Esta característica permite aumentar la posibilidad de mitigar el riesgo en el que puede incurrir un país en lo relativo al lavado de activos y financiación del terrorismo dentro de su sistema financiero.

A pesar de lo anterior, estas metodologías son incapaces de abarcar todo lo que significa mitigar e identificar los posibles riesgos LA/FT existentes en la economía de un país. Es decir, el monitoreo estadístico de los comportamientos financieros de los clientes que pueden incurrir en este tipo de actividades delictivas no es suficiente para determinar cuáles son los perfiles financieros que conducen a la ejecución de actividades criminales. Para complementar la parte estadística de estos controles, es necesario que los procedimientos requeridos por la normatividad colombiana sobre esta materia no dejen de lado aspectos tan importantes como lo son el contexto socio-económico y la coyuntura específica de cada sector de la economía. Los cuales podrían otorgar un mayor grado de certeza a los resultados que sirven para identificar movimientos sospechosos.

Para el caso de estudio, esta combinación metodológica entre técnicas de minería de datos y análisis económico sectorial reflejó resultados positivos según la perspectiva de cumplir el objetivo de identificar movimientos inusuales con un mayor grado de especificidad y discriminación según la

naturaleza de sus acciones. Es decir, en la medida que se entienda que las alertas generadas bajo los requerimientos actuales de la norma adquirieron un contexto capaz de establecer de manera jerárquica, cuales movimientos representaban un mayor riesgo de cometer crímenes financieros y cuales podrían sustentarse en la coyuntura económica puntual del sector de la economía al que pertenecen; la propuesta metodológica de este escrito logra plantear una manera eficaz para analizar los riesgos LA/FT asociados a las transacciones realizadas desde la entidad financiera estudiada hacia el mercado financiero en general. Y con esto, ayudar a que el objetivo del gobierno de prevenir e identificar el lavado de activos y la financiación del terrorismo que puedan ser cometidas sobre las operaciones realizadas dentro de su sistema financiero, se cumpla de una manera más eficiente y acertada.

Bibliografía

- ADRES. (OCTUBRE de 2017). *Administradora de los Recursos del Sistema General de Seguridad Social en Salud*. Obtenido de <http://www.adres.gov.co/BDUA/Estadistica-BDUA/Afiliados-por-Departamento-y-Estado-de-Afiliacion>
- AERONÁUTICA CIVIL. (OCTUBRE de 2017). *BASES DE DATOS HISTÓRICAS*. Obtenido de TRÁNSITO DE PASAJEROS POR ORIGEN-DESTINO NACIONAL: <http://www.aerocivil.gov.co/atencion/estadisticas-de-las-actividades-aeronauticas/bases-de-datos>
- BAMLI, B. A.-M. (agosto de 2017). *Basel Anti-Money Laundering*. Obtenido de <https://index.baselgovernance.org/ranking>
- Basel Institute on Governance. (2017). *Basel AML Index 2017*. Basel, Switzerland: International Center of Asset Recovery.
- Bolsa de Valores de Colombia. (OCTUBRE de 2017). *Información mercados locales*. Obtenido de Informes Bursátiles: <https://www.bvc.com.co/pps/tibco/portalbvc/Home/Mercados/informesbursatiles?action=dummy>
- Breiman , L., Friedman, J., Olshen, R., & Sontag, C. (1984). *Classification and Regression Trees*. *Wadsworth*.
- Brys, G., Hubert, M., & Struyf, A. (2004). *A robust measure of skewness*. Obtenido de <https://wis.kuleuven.be/stat/robust/papers/2004/medcouple.pdf>
- DANE. (OCTUBRE de 2017). *Encuesta Mensual de Comercio al por Menor y Comercio de Vehículos - EMCM*. Obtenido de <http://www.dane.gov.co/index.php/estadisticas-por-tema/comercio-interno/encuesta-emcm>
- DANE. (OCTUBRE de 2017). *Encuesta Mensual Manufacturera -EMM-*. Obtenido de <http://www.dane.gov.co/index.php/estadisticas-por-tema/industria/encuesta-mensual-manufacturera>
- DANE. (OCTUBRE de 2017). *Estadísticas de Licencias de Construcción -ELIC-*. Obtenido de <http://www.dane.gov.co/index.php/estadisticas-por-tema/construccion/licencias-de-construccion>

- Departamento Administrativo Nacional de Estadística (DANE). (2012). *Clasificación Industrial Internacional Uniforme de todas las actividades económicas. Revisión 4 adaptada para Colombia*. . Bogotá.
- Financial Action Task Force (GAFI). (2013). *National Money Laundering and Terrorist Financing Risk Assessment*.
- Financial Action Task Force (GAFI). (2016). *INTERNATIONAL STANDARDS ON COMBATING MONEY LAUNDERING AND THE FINANCING OF TERRORISM AND PROLIFERATION*.
- Ghanshyam, M. (1991). The Euclidean Distance Approach to Continue Utility Functions. *The Quarterly Journal of Economics*, Vol. 106, No. 3, 975 - 977.
- Gower, J. (1971). A General Coefficient of Similarity and Some of Its Properties. *Biometrics*, Vol. 27, No. 4., 857 - 871.
- Hubert, M., & Vandervieren, E. (2007). *An adjusted Boxplot for Skewed Distributions*. Obtenido de <https://wis.kuleuven.be/stat/robust/papers/2008/adjboxplot-revision.pdf>
- Jhonny R. Demey, L. P.-V. (2011). *Medidas de Distancia y de Similitud*.
- Kaufman, L., & Rousseeuw, P. (2005). *Finding Groups in Data*. Hoboken, New Jersey: Wiley & Sons.
- Levi, M., & Reuter, P. (2006). Money Laundering. *Crime and Justice*, Vol. 34, No. 1, 289 - 375.
- MacQueen, J. B. (1967). Some Methods for classification and Analysis of Multivariate Observations1. . *University of California Press*, 281 - 297.
- R., Q. (1993). *C4.5: Programs for Machine Learning*. Morgan Kaufmann Publishers. Obtenido de <http://www.rulequest.com/see5-unix.html>
- Richards, G., De la Iglesia, B., & Rayward-Smith, V. (2006). Clustering Rules: A Comparison of Partitioning and Hierarchical Clustering Algorithms. *Journal of Mathematical Modelling and Algorithms*, Vol. 5, 475 - 504.
- Santamaría Ruiz, W. (2017). *Técnicas de Minería de Datos para la Detección de Lavado de Activos*.
- Schneider, F., & Hametner, B. (2007). The Shadow Economy in Colombia: Size and Effects on Economic Growth. *Working Paper No. 0703. Department of Economics. Johannes Kepler University of Linz*.
- Sudjianto, A., Ming, Y., Kern, D., Nai, S., Zhang, A., & Cela-Díaz, F. (2010). Statistical Methods for Fighting Financial Crimes. *Technometrics*, Vol. 52, No. 1, 5 - 19.
- UIAF. (2014). *La Dimensión Económica del Lavado de Activos*. Bogotá: UIAF, Segunda Edición.
- UIAF, U. (2014). *Técnicas de minería de datos para la prevención del Lavado de Activos y La Financiación del terrorismo (LA/FT)*. García Solano y Compañía SAS.
- Universidad del Rosario, MinJusticia & UIAF. (2016). *Evaluación Nacional del Riesgo de Lavado de Activos y Financiación del Terrorismo*.
- Van der Maaten, L., & Hinton, G. (2008). Visualizing High-Dimensional Data Using t-SNE. *Journal of Machine Learning Research*, 2579 - 2605.

Apéndice

Anexo 1. Reglas de clasificación de segmentos personas naturales.

NOMBRE SEGMENTO	EDAD	OCUPACIÓN	CLIENTES
SEGMENTOS NATURALES 1*		INDEPENDIENTE	3196
SEGMENTOS NATURALES 2	<=6 AÑOS	SIN DEFINIR	18
SEGMENTOS NATURALES 3		JUBILADO, PENSIONADO	385
SEGMENTOS NATURALES 4	>69 AÑOS	EMPLEADO-SOCIO, RELIGIOSO	12
SEGMENTOS NATURALES 5		AMA DE CASA	197
SEGMENTOS NATURALES 6	>62 AÑOS, <= 69 AÑOS	EMPLEADO-SOCIO, RELIGIOSO	8
SEGMENTOS NATURALES 7		RENTISTA DE CAPITAL	431
SEGMENTOS NATURALES 8	>53 AÑOS, <= 62 AÑOS	EMPLEADO-SOCIO, SIN DEFINIR	20
SEGMENTOS NATURALES 9*		EMPLEADO	1558
SEGMENTOS NATURALES 10	>32 AÑOS, <= 53 AÑOS	EMPLEADO-SOCIO, RELIGIOSO, SIN DEFINIR	37
SEGMENTOS NATURALES 11		ESTUDIANTE	450
SEGMENTOS NATURALES 12	>6 AÑOS, <= 32 AÑOS	EMPLEADO-SOCIO, SIN DEFINIR	7

*En algunas circunstancias los segmentos otorgados por las reglas de segmentación podrían llegar a tener un número muy grande de integrantes, esto dificultaría realizar de una manera eficaz el procedimiento de definición de alertas. En función de generar unas alertas eficaces y dirigidas a segmentos específicos, se opta por fraccionar los dos segmentos más grandes de la categoría clientes. Este fraccionamiento se hace utilizando la variable patrimonio, de la cual se calculan los cuartiles¹⁶ para crear 4 nuevos sub-segmentos, compuestos de la siguiente manera:

Sub-segmento	Ocupación	Patrimonio
Segmentos Naturales 1.1	INDEPENDIENTE	\$0 – \$43'809.750
Segmentos Naturales 1.2	INDEPENDIENTE	\$43'809.750 - \$281'020.364
Segmentos Naturales 1.3	INDEPENDIENTE	\$281'020.364 - \$853'168.250

¹⁶ Los cuartiles son los valores por los que se puede dividir un conjunto de observaciones en 4 partes porcentualmente iguales. Los números definidos por los cuartiles indican hasta que valor puede llegar un porcentaje de la muestra. Por ejemplo, si un segmento tiene en su patrimonio el cuartil 1 con valor de 5'000.000 significa que el 25% de las observaciones tienen un patrimonio igual o inferior a ese monto.

Segmentos Naturales 1.4	INDEPENDIENTE	MAYOR A \$853'168.250
Segmentos Naturales 9.1	EMPLEADO	\$0 – \$118'582.000
Segmentos Naturales 9.2	EMPLEADO	\$118'582.000 - \$371'171.000
Segmentos Naturales 9.3	EMPLEADO	\$371'170.000 – \$875'621.000
Segmentos Naturales 9.4	EMPLEADO	MAYOR A \$875'621.000

Anexo 2. Reglas de clasificación de segmentos personas jurídicas.

NOMBRE SEGMENTO	ACTIVIDAD ECONÓMICA CÓDIGO CIU	CLIENTES
SEGMENTO JURIDICOS 1	0112, 0113, 0126, 0140, 0161, 0162, 0170, 0510, 0722, 0910, 1312, 1392, 1399, 1511, 1630, 1690, 1709, 1811, 2029, 2030, 2100, 2219, 2310, 2322, 2394, 2399, 2410, 2910, 2920, 2930, 3091, 3110, 3311, 4010, 4220, 4312, 4321, 4330, 4520, 4521, 4610, 4631, 4632, 4665, 4719, 4732, 4742, 4752, 4761, 4772, 4773, 4775, 4782, 4799, 4930, 5011, 5012, 5111, 5121, 5210, 5224, 5320, 5530, 5612, 5911, 6130, 6190, 6201, 6209, 6421, 6614, 6721, 7120, 7320, 7420, 7710, 7740, 7912, 7990, 8010, 8291, 8413, 8511, 8522, 8622, 8692, 9006, 9102, 9329, 9820	91
SEGMENTO JURIDICOS 2	3511,4511,4923,6493	40
SEGMENTO JURIDICOS 3	0150, 6511, 6621, 9609	32
SEGMENTO JURIDICOS 4	6492	24
SEGMENTO JURIDICOS 5	6613	78
SEGMENTO JURIDICOS 6	4663, 4664, 4711, 5813, 6399, 6512, 6630, 8544, 8691, 8699, 9603	60
SEGMENTO JURIDICOS 7	0119, 1040, 1051, 1921, 2022, 2023, 2212, 2511, 2593, 2811, 2819, 3520, 3811, 4390, 4662, 4731, 4741, 4751, 4755, 4921, 6424, 6431, 6432, 7410, 7730, 8292, 8512, 8610, 8810, 9900	60
SEGMENTO JURIDICOS 8	0124, 0125, 0130, 2229, 2429, 2599, 4642, 4645, 4661, 5222, 5511, 6422, 6423, 7220, 7911, 8030, 8430, 8513, 8621, 90, 9411	63
SEGMENTO JURIDICOS 9	0-SIN DEFINIR, 4210	22
SEGMENTO JURIDICOS 10	8299	30
SEGMENTO JURIDICOS 11	7490	100
SEGMENTO JURIDICOS 12	9491	27
SEGMENTO JURIDICOS 13	6810	159
SEGMENTO JURIDICOS 14	0141, 6820	30
SEGMENTO JURIDICOS 15	0210, 1089, 2395, 3290, 4620, 4641, 4754, 5229, 5611, 6020, 8543, 8559, 8560	52
SEGMENTO JURIDICOS 16	4771, 6612, 7310, 8211, 8530, 8790, 8890	42

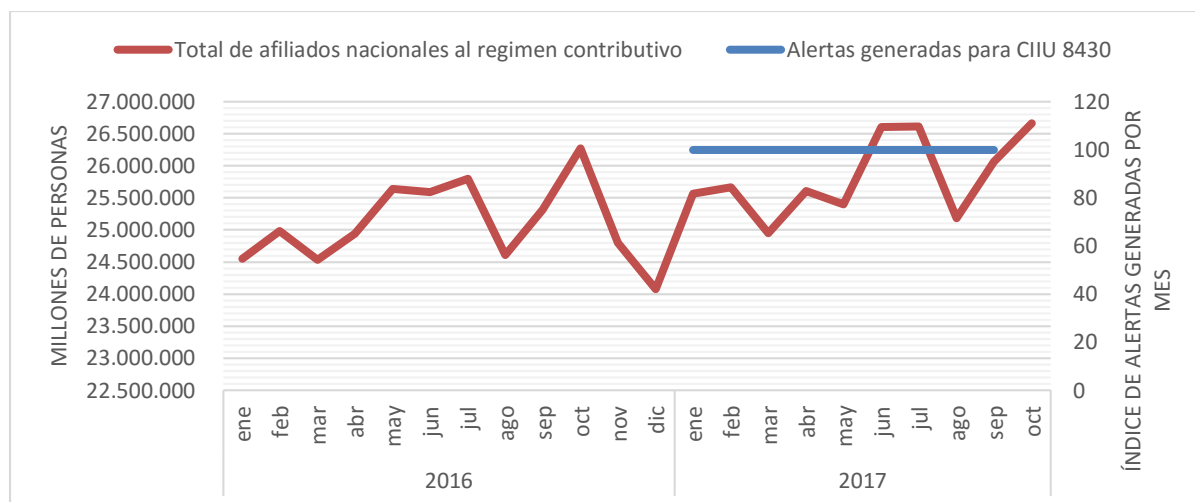
SEGMENTO JURIDICOS 17	4111	34
SEGMENTO JURIDICOS 18	1410	16
SEGMENTO JURIDICOS 19	0121, 4112, 4690	27
SEGMENTO JURIDICOS 20	4659, 4669	14
SEGMENTO JURIDICOS 21	6920, 7110	44
SEGMENTO JURIDICOS 22	9499	54
SEGMENTO JURIDICOS 23	4530	23
SEGMENTO JURIDICOS 24	4290	14
SEGMENTO JURIDICOS 25	7010	136
SEGMENTO JURIDICOS 26	6499	38
SEGMENTO JURIDICOS 27	7020	33
SEGMENTO JURIDICOS 28	6619	26
SEGMENTO JURIDICOS 29	6494	47
SEGMENTO JURIDICOS 30	6910	12

Anexo 3. Coeficientes de Correlación CIU vs Comportamiento de su mercado específico

Código CIU	Coeficiente de correlación Mercado-CIU
CIU 6810	0,22069302
CIU 5111	-0,44505066
CIU 6613	-0,26175927
CIU 8430	N/A*
CIU 3091	N/A*
CIU 4511	N/A*

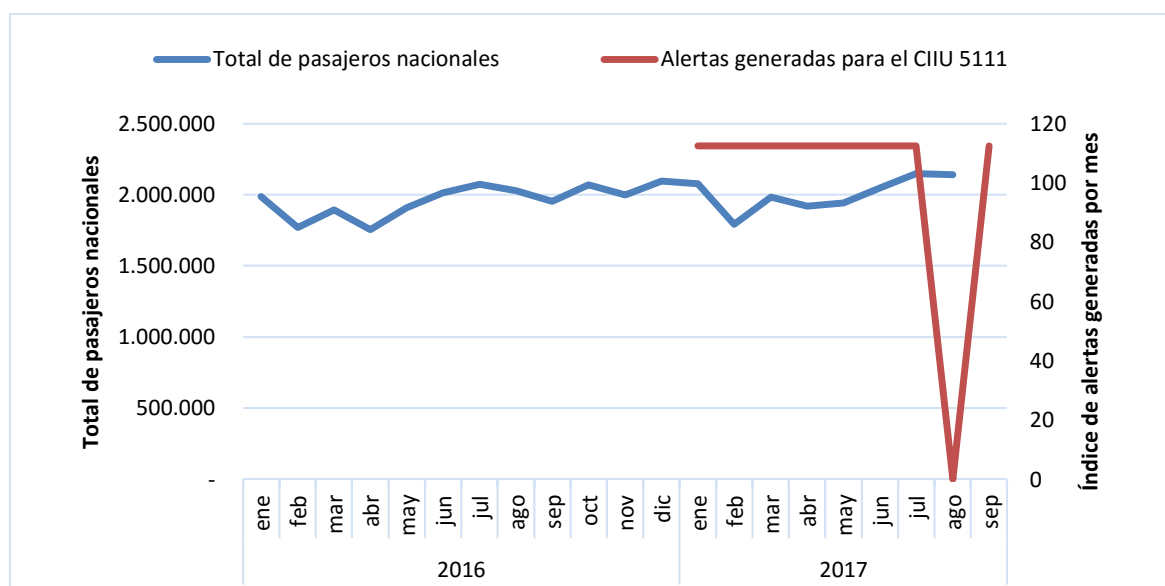
*Para estos códigos en específico no fue posible hallar un coeficiente de correlación entre las series de tiempo de las alertas generadas por códigos CIU y las series que representan el movimiento del mercado, en la medida que se entienda que las alertas generadas durante el periodo de estudio fueron totalmente estables de una alerta por mes, lo que no permitía normalizar el valor de su índice ya que la varianza era igual a 0.

Gráfico 11. Resultados código 8430 - actividades de planes de seguridad social de afiliación obligatoria vs total de afiliados al SSSI¹⁷ en el régimen contributivo.



Fuente: (ADRES, 2017). Cálculos propios.

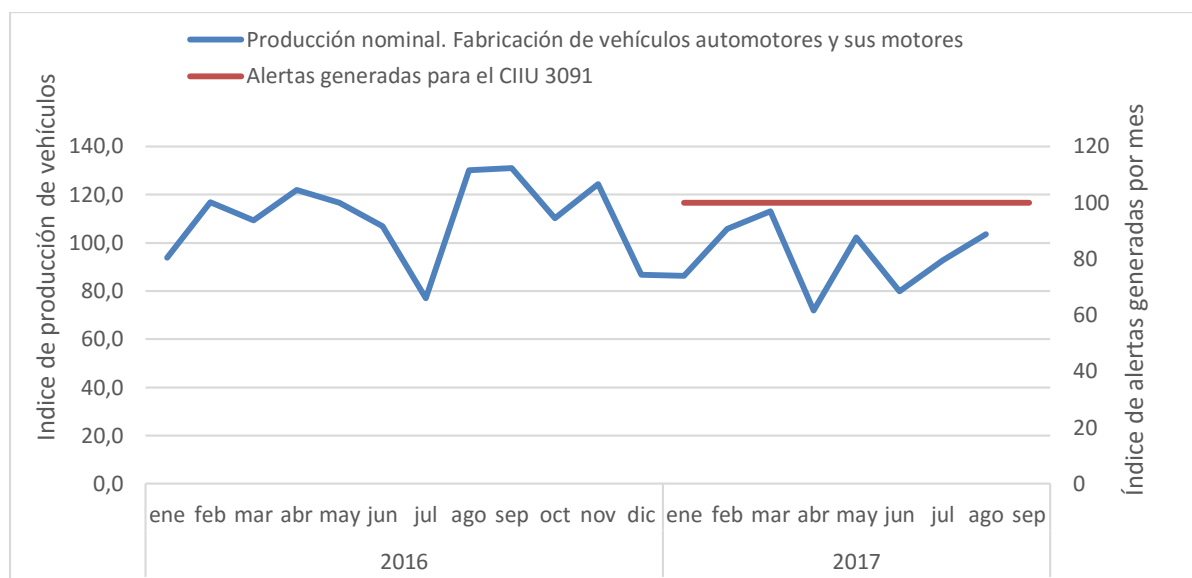
Gráfico 12. Resultados código 5111- transporte aéreo nacional de pasajeros



Fuente: (AERONÁUTICA CIVIL, 2017). Cálculos propios.

¹⁷ Sistema de Seguridad Social Integral

Gráfico 13. Resultados código 3091 - fabricación de motocicletas



Fuente: (DANE, 2017). Cálculos propios.